Targeted conservation genetics of the endangered chimpanzee

Peter Frandsen*[1,2,†], Claudia Fontsere*[3], Svend Vendelbo Nielsen[4], Kristian Hanghøj[2], Natalia Castejon-Fernandez[4], Esther Lizano[3], David Hughes[5,6], Jessica Hernandez-Rodriguez[3], Thorfinn Korneliussen[2], Frands Carlsen[1], Hans Redlef Siegismund[2], Thomas Mailund[4], Tomas Marques-Bonet[3,7,8,9], Christina Hvilsom[1].

1 Research and Conservation, Copenhagen Zoo, Roskildevej 38, 2000 Frederiksberg, Denmark.

2 Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark.

3 Institute of Evolutionary Biology, (UPF-CSIC), PRBB, Dr. Aiguader 88, 08003, Barcelona, Spain.

4 Bioinformatics Research Center, Department of Mathematics, Aarhus University, C. F. Møllers Allé 8, 8000 Aarhus C, Denmark.

5 MRC Integrative Epidemiology Unit at Universit of Bristol, Bristol, BS8 2BN, UK.

6 Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS8 2BN, UK.

7 Catalan Institution of Research and Advanced Studies (ICREA), Passeig de Lluís Companys 23, 08010, Barcelona, Spain.

8 CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri I Reixac 4, 08028 Barcelona, Spain.

9 Institut Català de Paleontologia Miquel Crusafant, Universitat Autònoma de Barcelona, Edifici ICTA-ICP, c/ Columnes s/n, 08193 Cerdanyala del Vallès, Barcelona, Spain.

Running title: Targeted conservation genetics

* These authors contributed equally to this work

† Corresponding author:
Peter Frandsen, PhD
Research and Conservation
Copenhagen Zoo
Roskildevej 38
2000 Frederiksberg, Denmark
E-mail: pef@zoo.dk

Running title

Targeted conservation genetics

Keywords

Conservation genetics, *ex situ* breeding programmes, illegal wildlife trade, *in situ* conservation, non-invasive sampling.

Word count

6 793

Running title: Targeted conservation genetics

Abstract

Populations of the common chimpanzee (*Pan troglodytes*) are in an impending risk of going extinct in the wild as a consequence of damaging anthropogenic impact on their natural habitat and illegal pet and bushmeat trade. Conservation management programmes for the chimpanzee have been established outside their natural range (*ex situ*), and chimpanzees from these programmes could potentially be used to supplement future conservation initiatives in the wild (*in situ*). However, these programmes have often suffered from inadequate information about the geographical origin and subspecies ancestry of the founders. Here, we present a newly designed capture array with ~60 000 ancestry informative markers used to infer ancestry of individual chimpanzees in *ex situ* populations and determine geographical origin of confiscated sanctuary individuals. From a test panel of 167 chimpanzees with unknown origins or subspecies labels, we identify 90 suitable non-admixed individuals in the European Association of Zoos and Aquaria (EAZA) *Ex situ* Programme (EEP). Importantly, another 46 individuals have been identified with admixed subspecies ancestries, which therefore over time, should be naturally phased out of the breeding populations. With potential for future re-introduction to the wild, we determine the geographical origin of 31 individuals that were confiscated from the illegal trade and demonstrate the promises of using non-invasive sampling in future conservation action plans. Collectively, our genomic approach provides an exemplar for *ex situ* management of endangered species and offers an efficient tool in future *in situ* efforts to combat the illegal wildlife trade.

Introduction

In an era of human-induced acceleration of species loss, often referred to as the sixth mass extinction era (Ceballos *et al.*, 2015), conservation efforts to save endangered species are calling for novel approaches to mitigate the ongoing extinction crisis.

Since the discovery of the common chimpanzee (*Pan troglodytes*), humans have been drawn to this charismatic species. Despite our fascination, human activities have led to a drastic decline in the population size of the chimpanzee. In the last two decades, chimpanzees have been listed as 'Endangered' at the species level in the IUCN Red List, with one of the four recognized subspecies, the western chimpanzee (*P. t. verus*) being listed as 'Critically Endangered' in the latest assessment (Humle *et al.*, 2016). Human encroachment on the natural range of the chimpanzee has further caused an intensified conflict between humans and chimpanzees (Hockings *et al.*, 2015). One by-product of the human wildlife conflicts has been a rise in opportunistic trafficking of chimpanzees, which, in recent years has become more organized and systematic (Stiles *et al.*, 2013). Besides wildlife trade, other continuous threats including habitat destruction, poaching for local consumption, and human linked disease outbreaks has led to a drastic decline in the wild chimpanzee populations (Humle *et al.*, 2016). Together, these threats emphasize the importance of a '*One Plan Approach*' conservation programme linking *in situ* and *ex situ* efforts (Traylor-Holzer *et al.*, 2019) to prevent the predicted extinction of chimpanzees within the current century (Estrada *et al.*, 2017).

Outside Africa, several regional chimpanzee conservation programmes exist, with the largest being the European Association of Zoos and Aquaria (EAZA) *Ex situ* Programme (henceforth EEP). The EEP targets the subspecies level and today, breeding programmes for two of the four recognized subspecies, the western chimpanzee (*P. t. verus*) and the central

chimpanzee (*P. t. troglodytes*) have been established (Carlsen and de Jongh, 2018) The primary aim of the EEP is to safeguard the survival of healthy self-sustaining populations targeting the taxonomical level of subspecies (Carlsen and de Jongh, 2018). The extant EEP populations consist of wild founders and descendants thereof. However, in times before high resolution genetic technologies were available and even in its early development, knowledge of subspecies labels and relatedness between founders were inaccurate and has led to admixture of subspecies in the captive population (Hvilsom *et al.*, 2013). Early attempts to add a genetic layer to the EEP management has confirmed that knowledge of subspecies ancestries, inbreeding and relatedness estimates are instrumental to preserve genetic diversity in captive populations (Hvilsom *et al.*, 2013). Yet, most recent attempts based on microsatellite markers (Hvilsom *et al.*, 2013), did not have the necessary resolution or predictive power to disentangle several generations of hybridizations in the EEP breeding population. Although we still do not know its full extent, hybridization between neighbouring subspecies of chimpanzees has been shown to occur in the wild (Hvilsom *et al.*, 2013; Prado-Martinez *et al.*, 2013; de Manuel *et al.*, 2016) and therefore, it is not unlikely that some founders in the EEP harbour shared ancestries from more than one subspecies. The current strategy in the EEP targets un-admixed breeding individuals and with the current methods, it is impossible to tell if small admixture proportions arose from an early *ex situ* hybridization event followed by several generations of backcrossing or from a naturally admixed founder. Therefore, founders are potentially being wrongfully excluded from the breeding programme due to their admixed ancestry.

The scenario outlined above, is by no means exclusive to captive management of chimpanzees but extends to practically any *ex situ* management programme of populations based on wild born founders with a taxonomical subdivision. When morphology alone is insufficient in taxonomical delimitation between subspecies or the targeted conservation units,

genetic resources becomes increasingly important. Yet, the choice of genetic resource is not always trivial. In response to a growing availability of different types of genetic resources with widely different applications, several studies have tried to develop guidelines based on the management requirements (see e.g. Grueber *et al.*, 2019; Norman *et al.*, 2019).

As described, the complexities in EEP management of chimpanzees requires a new rigorous solution as previous attempts using either mitochondrial DNA, or microsatellites have proven insufficient. With a genome-wide set of ancestry informative markers, we predict that it will be possible to obtain the desired depth of predictive power to infer ancestries in the present and previous generations and classify individuals with shared ancestries as either descendants of admixed founders or *ex situ* hybrids. This could provide the foundation of a possible reassessment of the current management strategies under the EEP and in turn, allow for inclusion of wild born hybrids in the breeding programme if these are found to resemble the diversity of the species in the wild.

In their natural range, chimpanzees have become a commodity and organized illegal trade poses a serious threat to the species. Over the period from 2005-2011 a reported minimum of 643 chimpanzees were harvested from the wild for illegal trade activities (Stiles *et al.*, 2013). However, extrapolations suggest that twenty times as many individuals have become victims of the illegal wildlife trade in that relatively short time span (Stiles *et al.*, 2013). While most of the captured individuals are sold as bushmeat, a considerable number of mostly juvenile chimpanzees end up in the illegal pet trade. When conservation authorities confiscate illegally kept chimpanzees, they are placed at wildlife sanctuaries, often arbitrarily based on availability of space and proximity to the confiscation site. Whilst some of the rescued chimpanzees require specialised lifetime care, others may be successfully reintroduced into their natural habitats after extensive preparation (Beck *et al.*, 2007). For chimpanzees destined

to lifetime care, proper management planning requires knowledge about relatedness among sanctuary chimpanzees in order to set up family groups. In cases, were chimpanzees are suitable for reintroduction, knowledge of geographical origin is essential as several studies have shown lineage-specific adaptations in all four subspecies in their respective geographical ranges (e.g. Nye *et al.*, 2018). In the first complete geo-referenced genomic map of the chimpanzee, de Manuel *et al.* (2016) portrayed a strong correlation between geographical origin and genetic diversity, where the former can be inferred solely based on the latter. Employing genetic testing at the site of confiscation (e.g. airports, transport hubs) would enable conservation authorities to infer geographical origin of confiscated individuals and with time, strive to facilitate a return of these individuals to a protected area in the region where they were captured. Alternatively, confiscated chimpanzees can be sent to a neighbouring sanctuary with housing capacity, where specialized care and rehabilitation can be provided, and if possible, future reintroduction can be planned. Genetic testing at an early stage of confiscation also has the potential to understand and help break trafficking routes and enable CITES authorities to track and enforce law control in situations where chimpanzees are housed in disreputable zoos and entertainment facilities. However, to be a practical tool in conservation, the genetic test needs to maximise the inference accuracy, require very little investment, and pose as little risk to animal health as possible. These requirements limit our choice of applicable data types. With a novel SNP array design where the level of genetic information is only surpassed by costly whole genome sequencing, we argue that our approach constitutes the most cost-efficient option for conservation management in situations where funding is often scarce and demands for rigorous solutions are high.

Using a selected panel of 59 800 targeted ancestry informative markers, we demonstrate the ability to infer robust estimates of ancestry in several generations of the EEP

chimpanzee breeding population. We further show how this set of ancestry informative markers can be used to determine geographical origin of confiscated individuals and demonstrate how these methodologies can readily be applied to using non-invasive sampling. In combination, these methods harbour great potential for future global management plans for the chimpanzee and provides an important exemplar for management of endangered species in general.

Materials and Methods

Samples

A total of 179 chimpanzee samples were collected and analysed in the present study (Suppl. File S1 SequencingStatistics.xlsx). For the purpose of cross-validation between sequencing batches and to test our methodology on non-invasive hair sampling, a number of individuals were sequenced in duplicates and triplicates, which lead to 167 unique individuals. 136 from the EEP population housed in 47 different European zoos and primate rescue and rehabilitation centres (Table S2), and 31 from eight sanctuaries across Africa (Table S3). To form a reference panel, we complemented the genotypes of EEP and sanctuary chimpanzees with whole genome data from 58 geo-referenced wild-born chimpanzees, representing the four chimpanzee subspecies, and additionally, one known admixed individual (*Ptv-Donald*) and one known descendant of wild born individuals (*Ptv-Clint*) (Prado-Martinez *et al.*, 2013; de Manuel *et al.*, 2016).

DNA extraction and library preparation

DNA was extracted using a standard phenol-chloroform protocol. Samples were quantified with a Qubit 2.0 fluorometer, Qubit® dsDNA BR Assay Kit (Thermo Fisher Scientific). DNA library preparation was carried out in three batches. For the first batch (24 samples) and the second batch (63 samples), extracted DNA was sheared with a Covaris S2 ultrasonicator using the recommended fragmentation settings to obtain a 350 bp insert size. For the third batch (92 samples) DNA was sheared using the recommended settings of Covaris S2 to obtain 200 bp insert size. The first batch of 24 libraries (with 6 more samples not used in this study) were prepared using 1.5 μg of DNA and the TruSeq DNA HT Sample Prep Kit (Illumina), following

manufacturer's instructions and 14 cycles of PCR amplification. The second batch of 63 samples (with 17 more samples not used in this study) were processed using 500 ng of starting DNA and following the custom dual-indexed protocol described by Kircher *et al.* (2012) and 12 cycles of PCR were done for indexing and amplification. The remaining 92 samples (with 2 more samples not used in this study) were processed using 200 ng of starting DNA following the BEST protocol (Carøe *et al.*, 2018) with minor modifications (initial reaction volume was incremented up to 50 µl to accommodate a larger amount of starting DNA and 10 cycles of PCR amplification). For this third batch, we used inline barcoded short adapters with the same seven nucleotide barcodes at the P5 and P7 adapters. Clean-ups were done using homemade SPRI beads (Rohland and Reich, 2012). Libraries were eluted in 25 µl of ddH_2O and quantified with an Agilent 2100 Bioanalyzer using a DNA 1000 assay kit.

Target Capture Design

We performed a target capture enrichment experiment using baits synthesized by Agilent Technologies. We targeted 59 800 autosomal sites that were ancestry informative markers and designed using the panTro4 genome. Marker selection was done using published chimpanzee genomes (Prado-Martinez *et al.*, 2013) and by applying a sparse PCA method on 10 Mbp bins of the genomes (Lee *et al.*, 2012). Variant sites were then weighted to identify the most informative markers for the first two principal components (PCs) and 200 AIMs were extracted per segment. The genome was binned to have an unbiased and evenly distributed sampling of the genome and to have enough resolution to provide percent ancestry in highly admixed individuals.

For target enrichment hybridization, libraries were pooled equimolarly based on a library prep method to obtain a total of 19 pools (see Supporting Information for a detailed

description of the targeted enrichment hybridization). PCR amplification product was cleaned up using our homemade SPRI beads (Rohland and Reich, 2012). Each enriched sample was then quantified on a NanoDrop, BioAnalyzer and then sequenced.

Fastq filtering and mapping

Libraries were sequenced on five lanes of a HiSeq 2500 ultra-high-throughput sequencing system, one lane for 24 chimpanzee samples, two lanes for 63 chimpanzee samples and two lanes for the remaining 92 samples. Inline barcoded libraries captured in the same pool (92 from Batch 3) were de-multiplexed using Sabre software v. 1.0 (https://github.com/najoshi/sabre).

Prior to mapping, paired-end reads were filtered to remove PCR duplicates using FASTUNIQ v. 1.1 (Xu *et al.*, 2012) and adaptors (*Illuminaclip*) and low quality first five bases in a read (*Slidingwindow:5:20*) were trimmed using TRIMMOMATIC v. 0.36 (Bolger *et al.*, 2014). Overlapping reads were merged with a minimum overlap of 10 bp and minimum length of final read to 50 bp, using PEAR v. 0.9.6 (Zhang *et al.*, 2014). Then, reads were mapped using BWA v. 0.7.12 (Li and Durbin, 2009) to the Hg19 reference genome (GRCh37, Feb.2009 (GCA_000001405.1)). PCR Duplicates were removed using PICARDTOOLS v. 1.95 (http://broadinstitute.github.io/picard/) with the *MarkDuplicates* option. Further filtering of the reads was done to discard secondary alignments and reads with mapping quality lower than 30 using SAMTOOLS v. 1.5 (Li *et al.*, 2009). We then filtered for the targeted space (4 bp around the selected SNP) using BEDTOOLS intersect v. 2.16.2 (Quinlan and Hall, 2010).

The total aligned reads were calculated by dividing the number of uniquely mapped reads (the remaining reads after removing duplicates) by the number of production reads. The on-target aligned reads were calculated by dividing the target filtered reads by the production reads. Then, the total coverage was calculated by dividing aligned bases by the

length of the assembly (Hg19) and the target effective coverage dividing the on-target bases by the targeted genomic space. Finally, the enrichment factor of the capture performance was calculated by taking the ratio between the on-target reads by total mapped reads over the target size by genome size.

Variant calling

Variant discovery was performed using GATK '*Unified Genotyper*' (DePristo *et al.*, 2011) for each sample independently with the following parameters *-out_mode EMIT_ALL_SITES -stand_call_conf* 5.0 *-stand_emit_conf* 5.0 *-A BaseCounts -A GCContent -A RMSMappingQuality -A BaseQualityRankSumTest*. Genotypes from each sample were combined in a single VCF using GATK `*CombineVariants*' (DePristo *et al.*, 2011) with *-genotypeMergeOptions UNIQUIFY –excludeNonVariant* parameters. We also included the genotype information of available whole genome data of aforementioned 58 wild-born geo-referenced chimpanzees and *Ptv-Donald* and *Ptv-Clint* (Prado-Martinez *et al.*, 2013; de Manuel *et al.*, 2016). Unless differently stated in separate analysis, the variants with a depth of coverage less than 3, a quality score less than 30 (QUAL<30), minor allele frequency of 0.005 and a missingness rate of > 60 % were removed using VCFTOOLS v. 0.1.12 (Danecek *et al.*, 2011). We only kept the genotypes that were inside the target space by using the -bed option in VCFTOOLS v. 0.1.12 (Danecek *et al.*, 2011).

Ancestry inference and inbreeding

We inferred proportions of shared ancestries in two approaches. First, to detect underlying genetic structure with a reduction of the dimensionality in the data, we performed a principle component analysis (PCA) using EIGENSOFT v. 6.1.3. (Price *et al.*, 2006). All samples were

included without pruning of sites in linkage disequilibrium or minor allele frequencies, in order to avoid exclusion of fixed sites between populations. Analyses on shared ancestry in *ex situ* and sanctuary populations were done with reference to the genetic structure in the wild born individuals with ADMIXTURE v. 1.2 (Alexander *et al.*, 2009). To avoid any bias introduced from a joint analysis with related individuals, each of the 167 unique individuals from the EEP and sanctuary populations were analysed separately one by one against a reference panel of all wild born individuals. After applying a minor allele frequency filter (--maf 0.05) in PLINK v. 1.07 (Purcell *et al.*, 2007) to exclude sites polymorphic in only one individual, a set of 45 542 sites where kept for analysis. Each analysis of ADMIXTURE v. 1.2 (Alexander *et al.*, 2009) was iterated 100 times under an EM optimization algorithm and termination criteria of a log-likelihood increase of 10^{-5} between iterations. A value of K=4 was chosen to obtain clusters in line with the four recognized subspecies of chimpanzees. To assess convergence, the 100 iterations were evaluated to ensure that iterations did not differ by more than one log-likelihood value.

For each of the individuals with admixture coefficients >0.99, we applied NGSRELATEv2 (Hanghøj *et al.*, 2019) to estimate pairwise relatedness and individual inbreeding coefficients based on population allele frequencies from each of the inferred admixture clusters, after excluding minor allele frequencies (MAF) <0.05 (see Supplementary Information).

Hybrid classification

To further explore the ancestry sharing in the EEP and sanctuary individuals and to be able to differentiate shared ancestry originating from the founding individuals and EEP hybrids, we developed a hidden Markov model (available on GitHub

http://gihub.com/svendvn/ImmediateAncestry) to allow for an inference of the posterior proportion of ancestries in the three immediate previous generations. Additionally, we estimate where these immediate ancestors belong in the pedigree. For full documentation of the model, see Supplementary Information.

Re-assignment of geographical origin

We applied the methodology of ORIGEN (CRAN R package https://cran.r-project.org/web/packages/OriGen/index.html) as described by Rañola, Novembre, & Lange (2014), to re-assign the geographical origin of confiscated sanctuary individuals. We applied the *FitOriGenModelFindUnknowns* parameter to the 1 690 highest ranked informative markers to assign individual geographical origin onto the allele frequency surface, inferred from the wild born reference panel.

Non-invasive sampling

To test our targeted capture approach on non-invasively collected hair samples, we sequenced three individuals where we had both blood samples, whole genome reference data and hair samples. Hair samples were capture sequenced using the same methodology as described above for blood samples, except we added a pre-treatment step in the DNA extraction of hair samples to enhance lysis of keratin. Shared ancestry and geo-graphical origin was analysed as described above.

Results

Capture sequencing and variant calling

First we quantified and assessed the performance of our capture methodology in the selected targeted space. We wanted to ensure sufficient representation of the targeted genomic regions to reliably call the selected variants. In a total of five lanes of HiSeq2500 we obtained ~1 000 million production reads, and on average, each sample received five million reads. After removing PCR duplicates and considering only primary alignments with a mapping quality higher than 30, we obtained an average of 3.6 million mapped reads (74.31%) per sample (Suppl. File S1). The average effective target coverage on the 59,800 autosomal SNPs was 21.69 X with 12.91% of on-target reads (four base pairs around the targeted SNP, Suppl. File S1) which fulfilled our theoretical prediction of 20 X. In terms of capture performance, this last statistic is an underestimate since the full length of the capture bait is 120 base pairs and in this analysis, we only considered the four base pairs around the targeted SNP. Still, we considered it to be more accurate since it is the true space where the informative SNP falls. Lastly, to summarise the performance of the capture methodology, we computed the enrichment factor that relates the number of aligned reads on the target space divided by the production reads, with the size of the target space to the size of the whole genome. The resulting enrichment factor of 89.31 X reasserts the advantages of capture to ensure enough coverage for genotyping purposes (Suppl. File S1).

Considering all samples without overlap, we obtained a total of ~150 000 genotypes. However the average number of SNPs called per sample was 30 337 sites passing the filtering steps (MAF 0.05 and max-missing 0.6, after we excluded samples ‘12103’ and

'12349' due to low coverage). The maximum number of SNPs called in one individual was 51 952 and the minimum was 10 783 (Figure S1). Among the variation found in western chimpanzees, only a third of these were polymorphic in the western chimpanzee (Table S1), yet, of the 46 260 polymorphic sites, 15 738 were private in the western chimpanzee (Figure S2). For fixed sites, the western chimpanzee also had the highest number of private sites (Figure S2). Among the four subspecies, the eastern chimpanzee had the highest total number of polymorphic sites, followed by the central chimpanzee, Nigerian-Cameroon chimpanzee, and western chimpanzee, respectively (Table S1).

Population structure, ancestry, and inbreeding

The major axes of variance in EEP and sanctuary individuals were explored with a principal component analysis with reference to the panel of geo-referenced individuals with known subspecies label from Prado-Martinez *et al.* (2013) and de Manuel *et al.* (2016). The first principal component (PC1) explained 70.49 % of the variance in our data, separating the western chimpanzees from the three other subspecies in the reference panel (Figure 1B). With 16.53 % of explained variance, PC2 separated the Nigerian-Cameroon chimpanzee, central chimpanzee, and eastern chimpanzee.

The majority of the 167 tested individuals from the EEP and sanctuary populations, clustered with either of the four reference populations, while a minor part of the individuals scattered in between the defined populations (Figure 1B). The inferred ancestries from the ADMIXTURE analysis conveyed the same patterns of genetic population structure separating the geo-referenced individuals into four distinct clusters with varying degree of ancestry sharing between geographically neighbouring subspecies (Figure 1C). With this as a reference, we assigned the EEP and sanctuary individuals into groupings in terms of their

ancestry patterns of either non-admixed or hybrids with multiple components of ancestry. Of the 167 tested individuals, 121 could be confidently assigned as non-admixed (admixture proportion from one subspecies ≥0.99). All 31 sanctuary individuals were assigned to subspecies level without evidence of admixture, where five clustered with the western chimpanzee, one with the Nigerian-Cameroon chimpanzee, one with the central chimpanzee, and 24 with the eastern chimpanzee. In the EEP population, we inferred the majority of the 90 non-admixed individuals to belong to the western chimpanzee (41), three with the Nigerian-Cameroon chimpanzee, 25 with the central chimpanzee, and 21 with the eastern chimpanzee. Of the remaining 46 EEP individuals, 38 were inferred to be hybrids with two ancestry components while the last eight had three ancestry components.

Of all the individuals from the EEP, sanctuary, and the reference panel with admixture coefficients >0.99, relatedness estimates were low (Figure S3-S6) while we identified eight individuals with inbreeding coefficients above 0.2 (Figure 1D). Within these eight individuals, all four subspecies were represented, as were wild and captive born chimpanzees.

Hybrid classification

To explore ancestry patterns in the previous three generations, we ran our ancestry classification model going back k = 3 generations and visualized the number of loci each ancestor in generation k contributed to the ancestral informative part of the genome (see Supplementary Information). In general, our method correctly estimated the expected ancestries of our reference panel individuals (Figure 2A). Several eastern and Nigerian-Cameroonian chimpanzee individuals were estimated to contain substantial ancestry components from the mutually neighbouring central subspecies. The known hybrid *Ptv-Donald* (Prado-Martinez *et*

374 *al.*, 2013) was estimated by the method to be at least 1/8 central chimpanzee, yet the large
375 proportion of loci that were assigned to the central chimpanzee in the posterior distribution
376 might suggest that *Ptv-Donald* could be as much as 1/4 central chimpanzee.
377 Similar to the ancestries inferred with ADMIXTURE, our method classified a
378 large fraction of the EEP and sanctuary individuals to have ancestors from only one subspecies
379 in the last three generations (Figure 1C, Figure 2B, Figure 2C). In general, individuals inferred
380 to belong to the eastern chimpanzee had third generation ancestors of central chimpanzee
381 ancestry (Figure 2B, Figure 2C). Similarly, four inferred central chimpanzees in the EEP
382 population, showed small proportions of ancestry from the Nigeria-Cameroon chimpanzee.
383 Comparably, one sanctuary individual, *Edward*, was inferred here as a Nigeria-Cameroon
384 chimpanzee with a small proportion of central chimpanzee ancestry. However, performing
385 posterior correction by replacing the low central chimpanzee ancestor with another high
386 posterior Nigeria-Cameroon ancestor, would likely make a more accurate estimate. Among the
387 admixed EEP individuals, our model showed similar results to those obtained with
388 ADMIXTURE but as ancestry patterns became increasingly complex (more than two ancestral
389 subspecies) our inferred posterior proportions became increasingly uncertain (Figure 2B, Figure
390 S12). We further observed that in some cases, small deviating (possibly deep coalescing)
391 segments could have let the model to prefer configurations in the ancestry patterns to switch
392 halves (Figure 2C), while the correct configuration would probably be a simple case of
393 hybridization in the parent generation.
394
395 *Geo-localisation*
396 Based on an allele frequency surface map, built from our reference panel of wild born
397 individuals, we determined the geographical origin of all 31 sanctuary individuals. Generally,

the inferred probabilities of geographical origin gave accurate estimates (i.e. high probabilities assigned to just one or a few adjacent grid cells) for all sanctuary individuals (Figure 3). Also, all individuals assigned to the natural range of their inferred subspecies label. The majority of our tested sanctuary individuals belonged to the eastern chimpanzee where the geographical origins were inferred to six provinces along the eastern part of the natural range of the subspecies. Seven of the eastern individuals had low probability estimates divided over a cluster of adjacent grid cells, with the highest ranking cell assigned probability of less than 0.1. All five western chimpanzee individuals were assigned to the same grid cell in the eastern limits of their range. The single individual from the Nigeria-Cameroon chimpanzee was assigned to a locality in Cameroon while the one central chimpanzee was assigned to the coastal region of Gabon.

Non-invasive sampling

Expanding our targeted capture approach to non-invasively collected hair samples, corroborated the results obtained with blood samples. ADMIXTURE estimates converged to the same result in the two sample types for all tested individuals and geographical origin was assigned to the same locality between samples (Figure 4, Figure S13-17). Compared to the reference, ancestry estimates in our capture array approach did not always reveal the minor components of shared ancestries found when including all variant sites in the genome (Figure 4).

Discussion

As an exemplar for conservation genetics of endangered species, we have designed a novel capture array that targets identified ancestry informative markers across the genomes of 24 wild born chimpanzees (Prado et al., 2013) and the PanTro4 reference genome. Acknowledging that the selected ancestry markers were derived from a relatively limited set of genomes, which could potentially introduce an ascertainment bias towards specific subspecies, we confirmed that our design has the power to correctly identify the subspecies of an extended panel of newly sequenced chimpanzee genomes (de Manuel *et al.*, 2016) (Figure 1). Based on this proof of concept, we sequenced 167 chimpanzees from the EEP and sanctuary populations and analysed subspecies ancestries and geographical origin. We further show how this approach can be extended to non-invasive samples with robust results.

Ancestry of the ex situ population

In our test panel of 167 chimpanzees, 136 were from the EEP population housed at 47 European zoos and rehabilitation centres. Based on information on disembarkation or place of capture, we know that the majority of chimpanzees who founded the current EEP population came from West Africa. In accordance to this, a majority of the 90 non-admixed individuals could be assigned to the western chimpanzee (Figure 1C). Our findings confirm that for the western chimpanzee, early efforts of the EEP that sought to identify a core group of non-admixed western chimpanzees using mitochondrial DNA (Jepsen and Carlsen *unpublished*) and microsatellites (Hvilsom *et al.*, 2013), have been momentarily successful. Yet, using similar methodologies, previous attempts have only managed to identify a small group of central chimpanzees since the breeding effort for this subspecies was established (Carlsen and de

Jongh, 2018). Here, we identify 25 central chimpanzee individuals in the EEP population that show no evidence of shared ancestry with other subspecies (Figure 1C), and hence from a genetic viewpoint, would qualify as a suitable bolster to the current breeding population. Similarly, the 21 inferred non-admixed eastern chimpanzee individuals could form the crucial starting point from where a separate breeding effort could be established under the EEP. In contrast to this, of our tested 136 EEP individuals, only three could be assigned to the Nigerian-Cameroonian subspecies (Figure 1C) and in general, of the four subspecies, the Nigeria-Cameroon chimpanzee is by far the least represented in the EEP population (Carlsen and de Jongh, 2018). Yet, with our targeted capture approach, it will now be feasible to scan the remaining EEP population (~1 000 housed individuals) for additional non-admixed chimpanzee individuals in order to explore the possibilities of creating separate breeding populations for the two remaining subspecies.

Still, with a presumed small EEP population of eastern and Nigerian-Cameroonian chimpanzees, it might prove difficult to avoid inbreeding, although our estimates suggests, that high inbreeding coefficients are not exclusive to these particular subspecies. In fact, individuals with inbreeding coefficients in the range of 0.2-0.4 were found in each of the four subspecies and includes both wild and captive born individuals (Figure 1D). It is therefore difficult to establish whether the amount of inbreeding in EEP individuals are a consequence of breeding among closely related individuals or whether it stems from inbred founders. In a few cases, like individual '*14073*', we know from reliable pedigree information, that this individual is the offspring of two full-siblings (Carlsen and de Jongh, 2018). For the large majority of the EEP population, this knowledge is not available or are associated with high levels of uncertainties. Together with accurate ancestry inferences, genetically-based inbreeding

estimates will be of high importance in management of the breeding population as will other factors such as age, fecundity, behaviour, and housing capacities.

Of our 136 tested EEP individuals, 46 were inferred to be of hybrid origin (Figure 1C). In terms of distinguishing founder individuals with shared ancestry components (wild born hybrids) from *ex situ* hybrids, our ancestry analyses show that the majority of our inferred hybrids are between non-neighbouring populations in the wild (e. g. between the western chimpanzee and either of the three other subspecies) and are therefore most likely the result of hybridization in the EEP breeding population. From a management standpoint, these should eventually be phased out of the breeding programme. Yet, some known hybrids have been allowed to breed under the current management. This has been done with the purpose to maintain population numbers in an interim period while the populations reach their target size and also to allow experienced females to pass on up-bringing behaviour to young individuals in the housed groups. To explore the extent of wild born hybrids in the EEP and the possibility of including these in the breeding efforts, we developed a new method for hybrid classification that can trace ancestry patterns three generations back. This could possibly allow us to distinguish between hybrids bred in captivity and wild born hybrids, where the latter could be included in breeding programmes, as they represent natural processes in the wild. However, two key requirements to such an inclusion are a better understanding of the extent of hybridization in the wild and an EEP management decision on what a suitable admixture threshold would be.

As validation for the hybrid classification model (see also Supplementary. Information), our method infers the known hybrid background of *Ptv-Donald* to have received at least 12.5 percent of its ancestry from the central chimpanzee, which is in the range of what was previously estimated using whole genome sequencing data (Prado-Martinez *et al.*, 2013).

Yet, in the EEP population, only a few of the inferred hybrids fit with the expectations of ancestry patterns in wild born hybrids. The majority of the inferred hybrids include a western chimpanzee ancestry component (Figure 2B), which is highly unlikely to occur in the wild due to the vast geographical distance to any neighbouring subspecies (Figure 1A). Of the eight inferred hybrids with adjacent distribution ranges, one central/Nigerian-Cameroonian and seven central/eastern hybrids (Figure 2B), we know from studbook information that all eight individuals were captive born (Carlsen and de Jongh, 2018) (Table S2). The only cases where our model might have picked up remnants from natural hybridizations are the ancestry components of central chimpanzee in what we inferred to be non-admixed eastern chimpanzees using ADMIXTURE (Figure 1C, Figure 2B). However, this could likely be due to a general limitation of our model to separate these two subspecies due to their evolutionary close relationship and history of allele sharing (Prado-Martinez *et al*., 2013; de Manuel *et al.*, 2016). Although we did not identify any wild born hybrids in the tested set of individuals, our model predictions will be highly useful in terms of pinpointing the timing of admixture and help to illuminate blanks in the studbook regarding possible sires.

Sanctuary ancestry and geographical origin

In contrast to the predominance of western chimpanzee individuals in the EEP population, the majority of the tested sanctuary individuals are inferred to belong to the eastern chimpanzee. Of the 31 tested individuals, we only find four that can be assigned to the western chimpanzee and a single individual from each of the Nigeria-Cameroon chimpanzee and the central chimpanzee (Figure 1C). When exploring ancestry patterns in the last three generations, we obtained similar results as in the EEP population, where small posterior proportions of central chimpanzee were found in individuals of the eastern chimpanzee (Figure 2C). This is most

likely due to the limitations of our model when it comes to distinguishing shared alleles between these two subspecies, and we do not infer any geographical origin close to possible contact zones between the two subspecies (Figure 3).

For western and Nigerian-Cameroonian chimpanzees, we obtained high probabilities in the assigned origins but with little spatial resolution. Essentially, all five western chimpanzee individuals assign to the same grid cell. As de Manuel *et al*. (2016) have previously shown, population structure inferred in the western and Nigerian-Cameroonian populations, may not offer enough resolution to provide fine scale determination of geographical origin. To improve origin estimates in these populations, it is crucial to obtain a better representation of georeferenced samples across their distribution ranges. This has been achieved for most of the central and eastern chimpanzee ranges, but with only one central chimpanzee individual (*Doris*), we cannot fully evaluate the prediction power and resolution for this subspecies. Nevertheless, the estimated geographical origin of *Doris* is very close to the reported confiscation site (Table S3), which gives us some assurance that future efforts to determine origins in the central chimpanzee will be possible. With a larger set of individuals from the eastern chimpanzee, we can start to appreciate the full potential of the method. The 24 analysed individuals can be assigned to geographical origins in six localities along the eastern edge of the distribution range of the eastern chimpanzee, where the majority originates from two locations in the northern and southern regions of the Democratic Republic of Congo (DRC) (Figure 3). First of all, this might tell us that these regions are heavily affected by poaching and illegal trafficking, although the abundance of confiscation sites might also be biased by the locality of contributing sanctuaries. Only further testing of individuals from sanctuaries across the species range will allow us to assess regional threat levels. However, with the inferred origins of the eastern chimpanzee individuals all along the eastern edge of the range, we can

conclude that the threats are not confined to only two regions for this subspecies but are distributed across the eastern boarders of the DRC.

When comparing the inferred geographical origins with the reported confiscation sites for all our tested sanctuary individuals (Table S3), it becomes apparent that the trafficking routes generally operate within a relatively local scale. Overall, we see that most of the tested individuals originate from locations that are within close proximity to where they have been confiscated, though with two notable exceptions, *Louise* and *Edward*. *Louise* was confiscated in Moscow, Russia and inferred to have originated from West Africa, while *Edward* was confiscated in Nairobi Airport, Kenya with inferred origin in Cameroon. This confirms that the illegal trade of wild chimpanzees spans beyond country borders and the African continent as reported in Stiles *et al.* (2013). Both individuals are now housed in sanctuaries where specialized care can be provided, yet, in these cases, both individuals have been placed in sanctuaries far from their geographical origin and possibly within mixed subspecies groups (other individuals from these sanctuaries have been assigned to different subspecies). Without proper knowledge of their ancestry, sanctuaries might face the same challenges as we have seen in the EEP population, with admixture of subspecies as a result of (unintended) breeding. Genetic testing at an early stage could help to ameliorate these challenges and as we have shown, our genomic approach extents to non-invasive sampling (Figure 4), making these methods both an accurate and practical tool in conservation efforts to help combat the illegal trade of chimpanzees.

We further predict that this approach will be self-empowering as sampling gaps in the distribution range of the chimpanzee are continuously covered and DNA extraction methods for non-invasive samples improve. This will significantly advance our predictive

power of geographical origin and provide valuable insight to shared ancestries in natural populations with positive knock-on effects to hybrid assessment in the *ex situ* populations.

Our capture array approach of targeting ancestry informative markers offers a standardized and cost-effective method that accurately guides *ex situ* and *in situ* conservation management programmes. At the current rate of decline, chimpanzees are predicted to go extinct within the current century (Estrada *et al.*, 2017). Conservation efforts might therefore, in a foreseeable future, be obligated to supplement wild populations with individuals from the *ex situ* populations as a last resort to prevent them from going extinct. Should it come to this, our approach facilitates the safeguarding of genetically self-sustainable populations that will have preserved a genetic profile that resembles their wild counterparts.

The current extinction crisis however, extends well beyond chimpanzees and the demand for molecular genetics to help guide future population management programmes is immense, ranging across the taxonomical scale of birds, reptiles, amphibians, and mammals. For the latter alone, more than ten EEP genetic projects are underway and globally, regional zoo associations are undertaking molecular genetic studies for which the present study serves as an important blueprint for linking *in situ* and *ex situ* conservation efforts.

Acknowledgements

The authors would like to thank all institutions who provided samples for this study: Stichting AAP Amersfoort Zoo, Antwerp Zoo, Zoo Delle Star ad Aprilia, Royal Burgers' Zoo Arnhem, Monde Sauvage Safari, Badoca Safari Park, Barcelona Zoo, Zoo Parc de Beauval, Bioparc Valencia, Borås Zoo, Parco Natura Viva, Edinburgh Zoo, Ölands Zoo, Plättli Zoo, Givskud Zoo/Zootopia, Safaripark Beekse Bergen, Hodonin Zoo, Xanthus Zoo, Kolmården Zoo, Krakow Zoo, Kristiansand Zoo, Lagos Zoo, Le Pal Zoo, Leipzig Zoo, Liberec Zoo, Lisbon Zoo, Madrid Zoo, Magdeburg Zoo, Olmense Zoo, Zoo Osnabrück, African Safari, Plaisance du Touch, Plzen Zoo, Centro de Rescate de Primates Rainfer, Zoological Center Tel Aviv - Ramat Gan, Safari Ravenna, Zoo di Roma, Leintal Zoo, Serengeti-park Hodenhagen, Reserve Africaine de Sigean, Wilhelma Zoo, Loro Parque Zoo, Touroparc, Twycross Zoo, AAP Primadomus, Warsaw Zoo, Schwabenpark, Zagreb Zoo, Cente International de Recherches Médicales de Franceville, Chimfunshi Wildlife Orphanage, Jeunes Animaux Confisques au Katanga, Ngamaba Island Chimpanzee Sanctuary, Tacugama Chimpanzee Sanctuary, The Chimpanzee Conservation Center. We further wish to thank Tom de Jongh and Lisbeth Borbye for valuable language editing and comments to the manuscript and Abigail Ramsøe for early developmental stages of the hybrid classification model.

PF is supported by the Innovation Fund Denmark doctoral fellowship programme and the Candys Foundation. CF is supported by “la Caixa” doctoral fellowship programme. TMB is supported by BFU2017-86471-P (MINECO/FEDER, UE), U01 MH106874 grant, Howard Hughes International Early Career, Obra Social "La Caixa" and Secretaria d’Universitats i Recerca and CERCA Programme del Departament d’Economia i Coneixement de la Generalitat de Catalunya (GRC 2017 SGR 880)

Conflict of Interest

The authors declare no conflicts of interest.

Data Archiving

Data will be archived at a publicly accessible repository (Dryad) as a VCF file containing all samples included in the analyses.

REFERENCES

Alexander DH, Novembre J, Lange K (2009). Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Res* **19**: 1655–1664.

Beck B, Walkup K, Rodriques M, Unwin S, Travis D, Stoinski T (2007). *Best Practice Guidelines for the Re-introduction of Great Apes*. Gland, Switzerland.

Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Carlsen F, de Jongh T (2018). *European Studbook for chimpanzee (Pan troglodytes)*. Copenhagen.

Carøe C, Gopalakrishnan S, Vinner L, Mak SST, Sinding MHS, Samaniego JA, *et al.* (2018). Single-tube library preparation for degraded DNA. *Methods Ecol Evol* **9**: 410–419.

Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Sci Adv* **1**: e1400253.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, *et al.* (2011). The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, *et al.* (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**: 491–8.

Estrada A, Garber PA, Rylands AB, Roos C, Fernandez-duque E, Fiore A Di, *et al.* (2017). Impending extinction crisis of the world's primates: Why primates matter. *Sci Adv* **3**.

Grueber CE, Fox S, McLennan EA, Gooley RM, Pemberton D, Hogg CJ, *et al.* (2019).

630 Complex problems need detailed solutions: Harnessing multiple data types to
631 inform genetic management in the wild. *Evol Appl* 12: 280–291.
632 Hanghøj K, Moltke I, Andersen PA, Manica A, Korneliussen TS (2019). Fast and
633 accurate relatedness estimation from high-throughput sequencing data in the
634 presence of inbreeding. 8: 1–9.
635 Hockings KJ, McLennan MR, Carvalho S, Ancrenaz M, Bobe R, Byrne RW, *et al.*
636 (2015). Apes in the Anthropocene: Flexibility and survival. *Trends Ecol Evol* 30:
637 215–222.
638 Humle T, Maisels F, Oates JF, Plumtre A, Williamson EA (2016). Pan troglodytes.
639 *IUCN Red List Threat Species*.
640 Hvilsom C, Frandsen P, Børsting C, Carlsen F, Sallé B, Simonsen BT, *et al.* (2013).
641 Understanding geographic origins and history of admixture among chimpanzees
642 in European zoos, with implications for future breeding programmes. *Heredity*
643 *(Edinb)* 110: 586–93.
644 IUCN (2015). IUCN Red List of Threatened Species. *Version 20153*:
645 www.iucnredlist.org.
646 Jepsen BI, Carlsen F Genetic identification of West African Chimpanzee, Pan
647 troglodytes verus, based on mitochondrial DNA analysis. *unpublished*.
648 Kircher M, Sawyer S, Meyer M (2012). Double indexing overcomes inaccuracies in
649 multiplex sequencing on the Illumina platform. *Nucleic Acids Res* 40.
650 Lee S, Epstein MP, Duncan R, Lin X (2012). Sparse principal component analysis for
651 identifying ancestry-informative markers in genome-wide association studies.
652 *Genet Epidemiol* 36: 293–302.
653 Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler

654 transform. *Bioinformatics* 25: 1754–1760.
655 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* (2009). The
656 Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.
657 de Manuel M, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J, *et al.*
658 (2016). Chimpanzee genomic diversity reveals ancient admixture with bonobos.
659 *Science* 354: 477–481.
660 Norman AJ, Putnam AS, Ivy JA (2019). Use of molecular data in zoo and aquarium
661 collection management: Benefits, challenges, and best practices. *Zoo Biol* 38:
662 106–118.
663 Nye J, Laayouni H, Kuhlwilm M, Mondal M, Marques-Bonet T, Bertranpetit J (2018).
664 Selection in the Introgressed Regions of the Chimpanzee Genome. *Genome Biol*
665 *Evol* 10: 1132–1138.
666 Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, *et al.*
667 (2013). Great ape genetic diversity and population history. *Nature* 499: 471–5.
668 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006).
669 Principal components analysis corrects for stratification in genome-wide
670 association studies. *Nat Genet* 38: 904–909.
671 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, *et al.* (2007).
672 PLINK: A tool set for whole-genome association and population-based linkage
673 analyses. *Am J Hum Genet* 81: 559–575.
674 QGIS (2018). QGIS Geographic Information System.
675 Quinlan AR, Hall IM (2010). BEDTools: a flexible suite of utilities for comparing
676 genomic features. *Bioinformatics* 26: 841–2.
677 Rañola JM, Novembre J, Lange K (2014). Fast spatial ancestry via flexible allele

678 frequency surfaces. *Bioinformatics* 30: 2915–22.
679 Rohland N, Reich D (2012). Cost-effective, high-throughput DNA sequencing libraries
680 for multiplexed target capture. *Genome Res* 22: 939–946.
681 Stiles D, Redmond I, Cress D, Nellemann C, Formo RK (2013). *Stolen Apes - The Illicit*
682 *Trade in Chimpanzees, Gorillas, Bonobos and Orangutans*. Birkeland Trykkeri AS,
683 Norway.
684 Traylor-Holzer K, Leus K, Bauman K (2019). Integrated Collection Assessment and
685 Planning (ICAP) workshop: Helping zoos move toward the One Plan Approach.
686 *Zoo Biol* 38: 95–105.
687 Xu H, Luo X, Qian J, Pang X, Song J, Qian G, *et al.* (2012). FastUniq: A Fast De Novo
688 Duplicates Removal Tool for Paired Short Reads. *PLoS One* 7.
689 Zhang J, Kobert K, Flouri T, Stamatakis A (2014). PEAR: a fast and accurate Illumina
690 Paired-End reAd mergeR. *Bioinformatics* 30: 614–620.
691

692 Figure 1

693 Subspecies ancestry in wild and captive populations of chimpanzees. A) Geographical
694 distribution ranges of the four chimpanzee subspecies (IUCN, 2015; QGIS, 2018). B)
695 Population structure by principal component decomposition of sanctuary and the EAZA Ex situ
696 Programme (EEP) populations with reference to wild born individuals. C) Shared ancestry
697 inferences of sanctuary and EEP individuals summarised from individual ADMIXTURE
698 analysis against the reference panel of wild born individuals. Individuals from the reference
699 panel are labelled with a subspecies ancestry prefix and known sample name in previous
700 literature (Prado-Martinez *et al.*, 2013; de Manuel *et al.*, 2016), sanctuary individuals are
701 labelled with common sample name identifiers, and individuals from the EEP are labelled by
702 studbook number (Table S2, Table S3). D) Inbreeding coefficients for all individuals with
703 admixture proportions >0.99 in either of the four inferred clusters. Clusters are colour labelled
704 in accordance to A, B, and C.
705

706 Figure 2

707 Hybrid ancestry in A) the reference panel, B) the EEP population, and C) the sanctuary
708 population. The estimated posterior ancestries, θ is shown for the eight ancestors $k = 3$
709 generations back in time, for each individual in the three populations. The ancestors are ordered
710 according to the "unphased" pedigree in the bottom of the plot. The width of each rectangle
711 indicate the expected proportion of loci that are assigned to that ancestor (conditioned on the
712 estimate of θ). Small widths suggest deviations from the model and features that could be
713 improved by posterior correction.
714

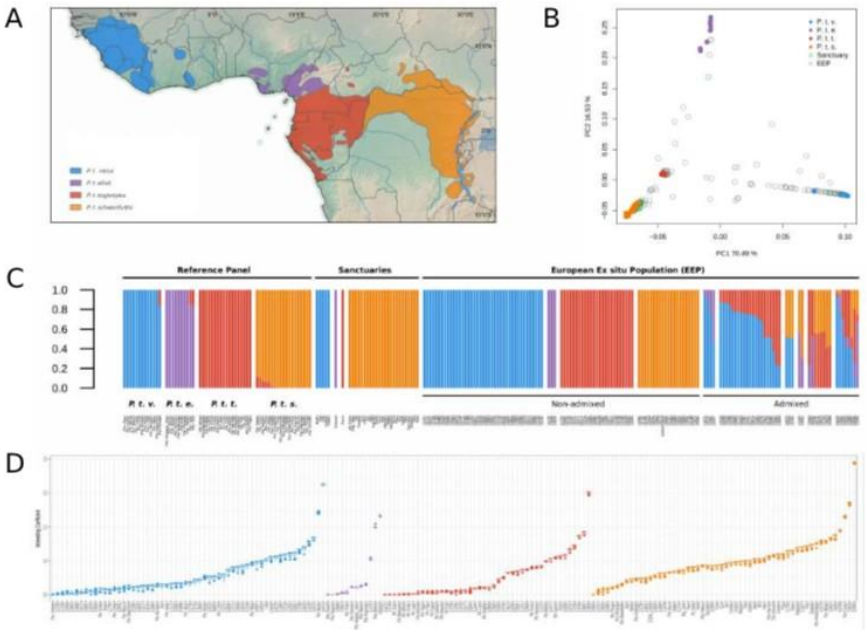
Figure 3

Geographical origin estimates for sanctuary individuals. Based on the allele frequency surface map of the reference panel, sanctuary individuals are assigned probabilities of geographical origin, here summarized from individual estimates.

Figure 4

Ancestry and geographical origin estimates from non-invasive samples. A) Geographical origin estimates from hair samples based on the allele frequency surface map of the reference panel, tested individuals are assigned probabilities of geographical origin, here summarized from individual estimates with comparison to blood samples (Figure S13-17). B) Shared ancestry estimates for hair samples compared to whole genome reference data and capture sequenced data from blood.

728 Figure 1

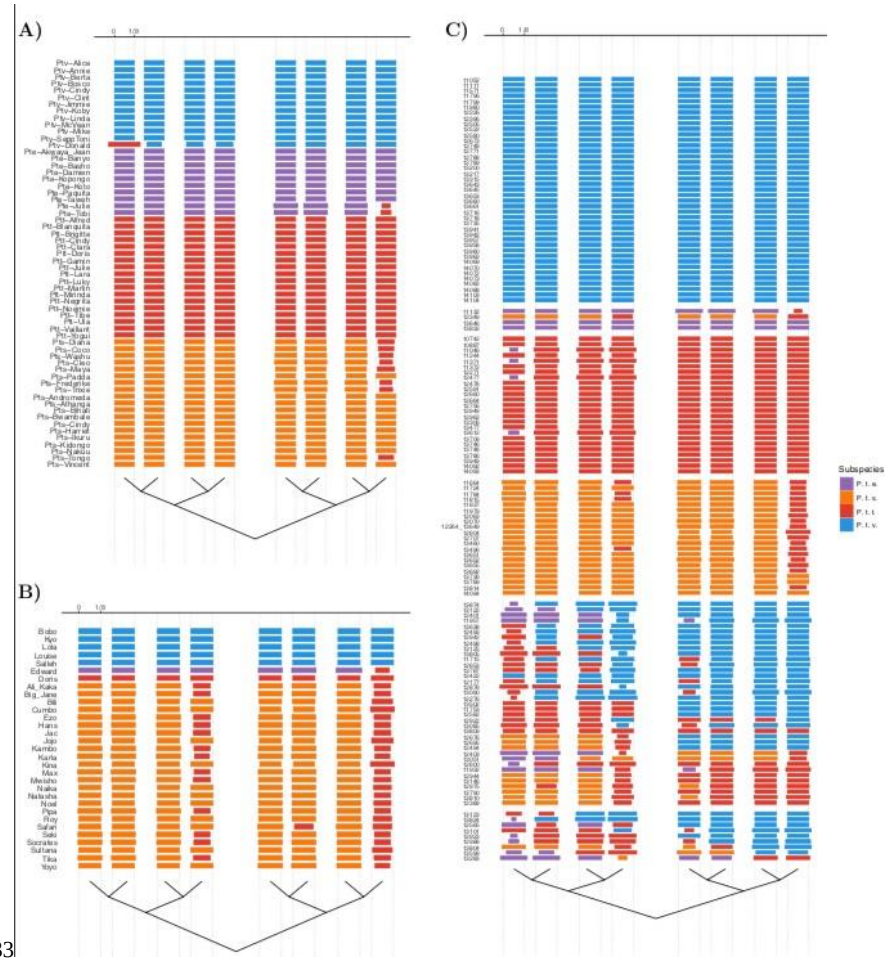


729

730

731

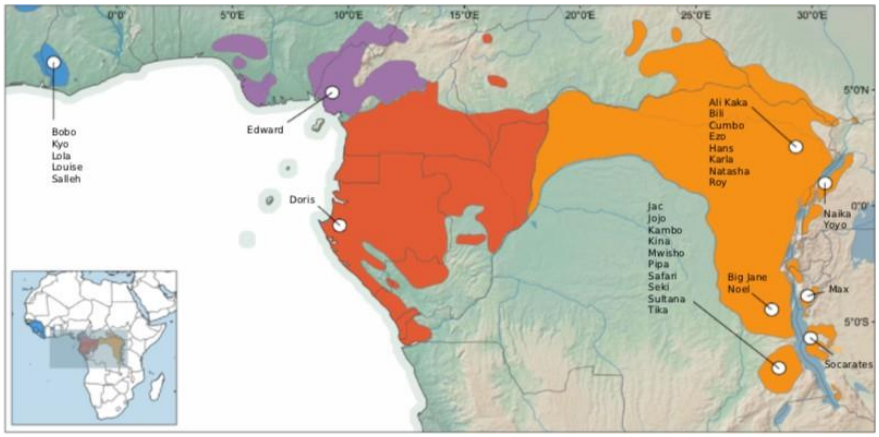
732 Figure 2



733

734

735 Figure 3

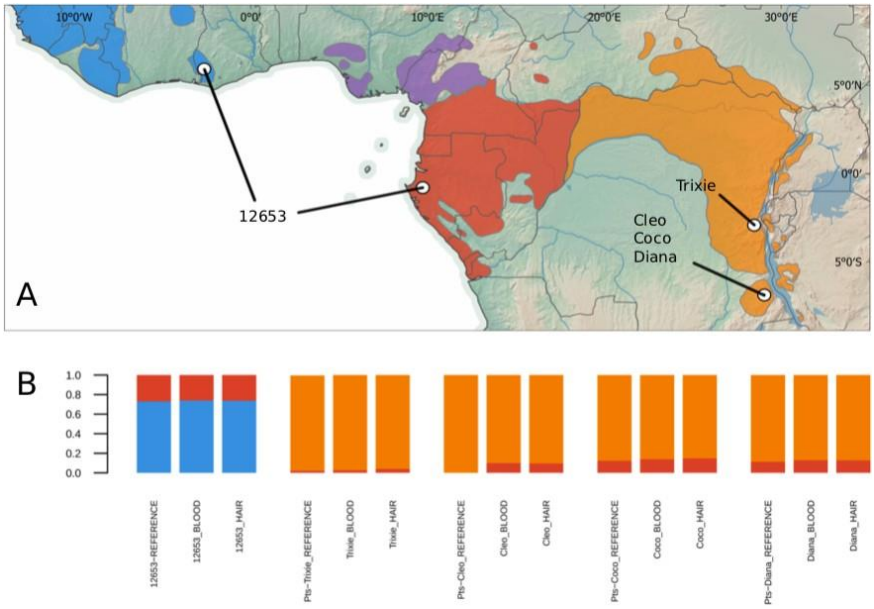


736

737

738

739 Figure 4



740